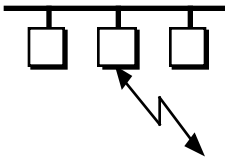

Networks and Communications Consulting

21885 Bear Creek Way
Los Gatos, CA 95030
(408) 395-5700
FAX (408) 395-1966
Internet: seifert@netcom.com

Technical Report

Issues in LAN Switching and Migration from a Shared LAN Environment

Author: Rich Seifert
November, 1995

TABLE OF CONTENTS

INTRODUCTION

Subject and Scope

Only a few years ago, Ethernet switching was considered leading edge, “exotic” technology. As the bandwidth needs of LAN-based applications has grown, users have begun to recognize the power of switching as a means of improving network performance in an incremental manner, without requiring a complete overhaul of the LAN infrastructure. Switching has emerged as one of the most powerful tools in the network designer’s arsenal.

This paper shows how the traditional approaches to LAN internetworking fail to adequately meet the needs of LAN application growth in many environments. Switching technology is shown to have superior behavior, and to provide better performance at lower cost than traditional routed internetworks. The limitations of switching are also discussed, so that users can best determine whether their network environment is amenable to using switching technology. We show how switching enables new capabilities in LANs that were not possible before, including: microsegmentation, full-duplex operation, and virtual LANs. Switching offers an easy way to integrate higher speed technologies, including Fast Ethernet and ATM, into growing networks. Finally, we consider a few important models of usage for switch deployment.

Reader Assumptions

This paper is not intended to be a primer on internetworking. The reader is assumed to be generally familiar with:

- Local Area Network technology, particularly Ethernet and 10BASE-T,
- LAN addressing concepts, including multicast and broadcast addressing,
- The basic operation and uses of bridges and routers.

THE CURRENT MODEL OF LAN INTERNETWORKING

Most practical networks today have an extent that far exceeds the possibility of locating all users and network resources on a single LAN. In most organizations:

- There are more users than can be accommodated by a single workgroup LAN,
- Users are geographically dispersed such that a single, flat network can not cover the required distances,
- Organizational boundaries must be reflected in the network design for administrative, management and budget reasons,
- There is more application demand than can be accommodated by a single network.

Figure 1 depicts the common model of an “enterprise-wide internetwork” which reflects these practical limitations.

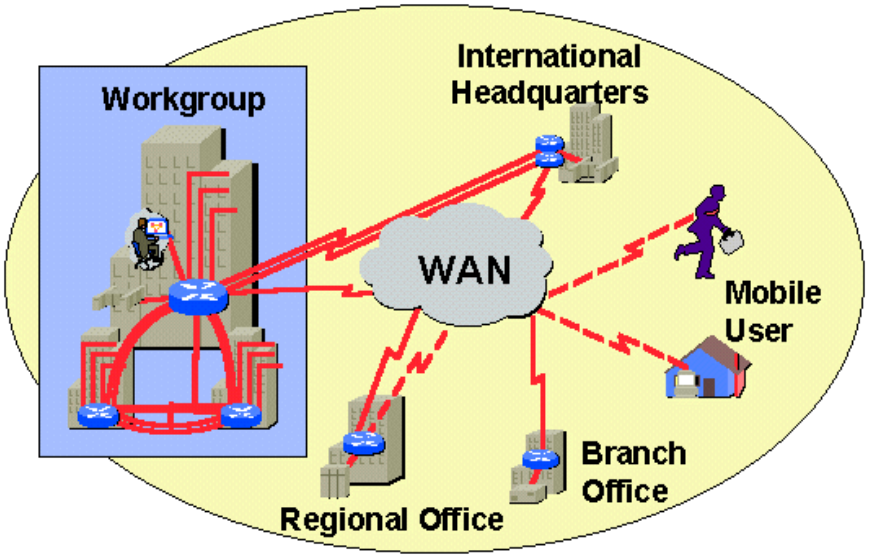

Figure 1: Current Model of Routed Enterprise Network

In this typical model:

- Workstations and servers are connected directly to LANs, along logical workgroup boundaries (i.e., users within a workgroup share the workgroup services on a single LAN). The most common workgroup LANs use Ethernet technology.
- Workgroup LANs are interconnected within a building or campus by routing the workgroup traffic to a backbone network. Depending on traffic requirements, the most common backbone networks use Ethernet or FDDI technology.
- Wide-area connectivity is provided across the enterprise by routing inter-site traffic across an enterprise backbone. Enterprise backbones use a wide range of technologies, including point-to-point links, frame relay, X.25, ATM, etc., depending on the performance requirements of the network.

Distributed and Collapsed Backbones

There are two common methods of providing a *backbone* to interconnect workgroup LANs. A backbone is a network whose primary purpose is the interconnection of other networks, as opposed to providing a direct attachment point for communicating devices. Regardless of whether routing or bridging is employed, a backbone may be either *distributed* or *collapsed*. In a distributed backbone (see Figure 2), the backbone network is brought to the internetworking devices. Geographically disperse internetworking devices are connected to the backbone, typically at wiring closets, to provide workgroup interconnectivity. In a collapsed backbone (see Figure 3), the backbone consists of a high-performance internetworking device. The workgroup networks must be “brought to” the backbone, often through point-to-point fiber links.

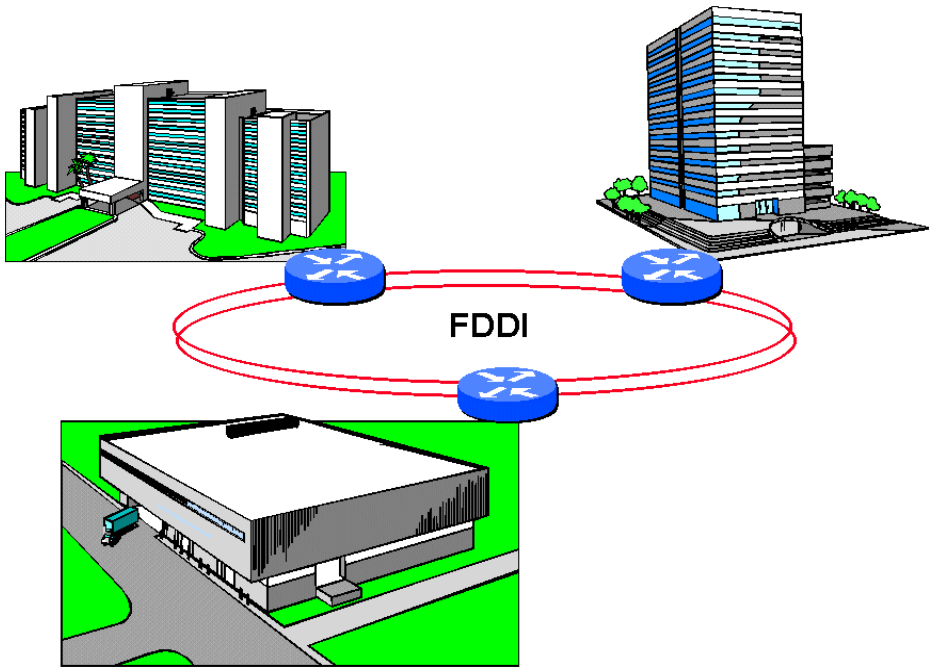

Figure 2: Distributed Backbone

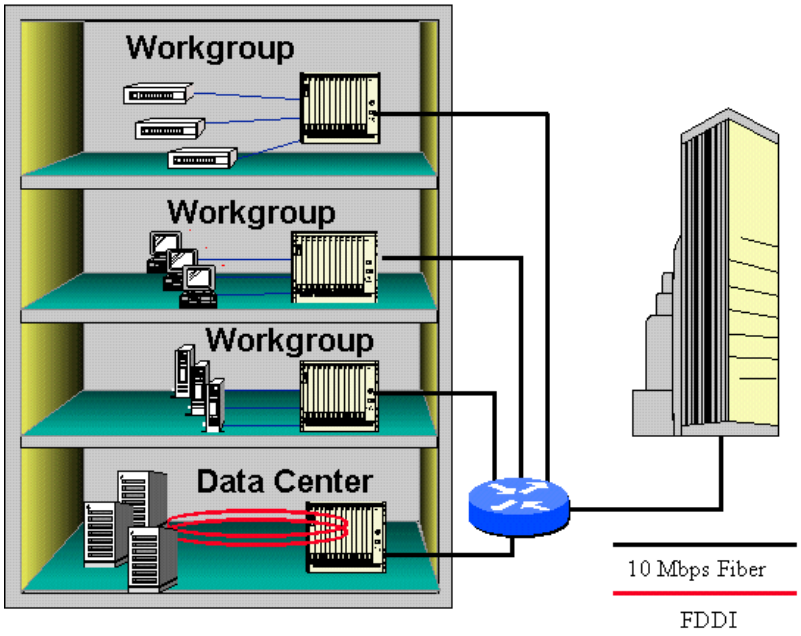

Figure 3: Collapsed Backbone

There are advantages and disadvantages to both approaches:

	Pro	Con
Distributed Backbone	Easily covers large distances Ease of expansion Robust (no single points of failure)	Lower total backbone bandwidth usually achievable
Collapsed Backbone	Lower cost per aggregate unit bandwidth Improved backbone security Single point of focus for network management	Single point of failure Impractical over very long distances (primarily suitable for campus)

Distributed backbones provide moderate-to-high bandwidth over moderate-to-long distances, albeit at relatively high cost. FDDI, a commonly used campus backbone, provides 100 Mbps of capacity over a 100 km ring circumference, but incurs a cost penalty. Typical high-performance FDDI interfaces cost thousands of dollars, plus the cost of installing the required fiber ring itself.

A collapsed backbone can provide hundreds or thousands of Mbps aggregate capacity at lower connect costs than FDDI, because the "backbone network" consists of the backplane of the internetworking device. By concentrating the internetworking functions into a single device, very high capacity can be provided at lower cost per Mbps. In addition, the backbone can be made more secure by simply limiting physical access to the collapsed backbone device.

The 80/20 Rule for Locating Resources

Perhaps the *most* important (and overlooked) network design issue is proper logical placement of network resources. In order to maximize network performance (and user satisfaction), workstations should be placed on the same logical network as the servers that they use most. This minimizes the need for high-performance internetwork devices and high capacity backbone networks.

A reasonable rule-of-thumb is that, in a properly designed environment, 80% of the traffic on a given network is local (destined for a target in the same workgroup); not more than 20% of traffic should require internetworking. Internetwork congestion may be an indication that traffic patterns are not obeying this 80/20 rule. It may be easier to improve application performance by either:

- Moving resources so that what was formerly internetwork traffic becomes localized to a workgroup,
- Moving users (logically, if not physically) so that the workgroups more closely reflect the traffic patterns, or
- Replicating resources (e.g., adding servers) so that users can access them locally without having to cross the internetwork.

It is important to remember that the demands placed on the internetworking infrastructure are a strong function of the local / remote traffic distribution.

Rationale for Inter-Workgroup Routing

Even given the cost and complexity of providing an internetworking backbone, there are several good reasons to use a traditional routed infrastructure:

- *Security*. Routers provide greater capabilities for network managers to implement security policies. Access control and prevention of unauthorized snooping are enhanced by the capabilities of network layer routing.
- *Network Management*. A network manager can exercise greater control over the behavior of an internetwork which uses traditional routing. Modern routers provide the means for fault isolation, performance monitoring, parametric control, routing policy implementation, etc.
- *Firewalls*. A routed internetwork provides a means for automatic containment of network problems. The failure of a single workgroup LAN, or the idiosyncrasies of a single application can be prevented from causing internetwork-wide catastrophic failure. Device or software faults causing high levels of congestion (including multicast congestion) can be contained to the smallest possible domain.

While these are all valid reasons to use a classical routed approach to internetworking, there are problems that arise from this paradigm that limit its effectiveness.

Issues Arising from the Routed Model

(1) *In a routed infrastructure, logical connectivity equates to physical connectivity*. Routers divide internetworks into workgroups or *subnetworks*. Subnetworks are associated on a quasi-static basis with a given port on a router. This is how the router makes the decision as to where to forward packets destined for a given subnetwork. If a user needs to be moved from one subnetwork to another, it is necessary to move that user's connectivity from one port of a router to another. In general, this implies a change in physical connectivity (e.g., at a patch panel in the wiring closet). As will be shown later in this paper, other paradigms allow the possibility of *dynamic* assignment of users to logical subnetwork, through the creation of *virtual LANs*.

(2) *Changing logical connectivity in a routed infrastructure may require reconfiguration of end stations.* Network layer protocols (e.g., IP) require that addresses, default routers, and other parameters be configured in each end station. This is normally a manual process. Reorganizing a set of stations into a new logical grouping may require that a network administrator manually change the configuration parameters in each station, individually. This is a tedious and time-consuming task, as well as being prone to data entry error.

In many cases, the cost of the moves and changes associated with re-defining subnetworks in a routed infrastructure is a significant component of the total cost of managing and maintaining the internetwork.

(3) *The performance of the internetwork may become router-limited.* In a routed backbone, all internetwork traffic must pass through at least one router. As internetwork traffic grows, the performance of the total system can quickly become limited by the performance of the routers themselves. Routers have greater capabilities and flexibility than bridges, but this implies that they have to do more processing on each packet. In addition to the critical task of determining where to forward each packet, routers typically must:

- Modify each forwarded packet for lifetime control (and possibly other functions as well) and recompute header checksums,
- Support the routing protocol that is dynamically updating the locally-held topology map,
- Implement an error reporting protocol (e.g., ICMP),
- Collect and organize management counters, statistics, and state information,
- Respond to management commands, etc.

All of these use processing power and reduce the ability of the router to handle heavy traffic loads. Obviously, all of these concerns can be overcome by increasing the capability of the router, but this comes at a price. At any given price, router performance can become the bottleneck in the internetwork.

(4) *Routed internetworks were not designed for bandwidth-intensive applications.* With traditional network applications such as E-mail, remote terminal access and simple file transfer, routers may be a perfectly reasonable solution. But applications are changing, and their network demands are growing. With text-based applications, file transfers were limited in size. With the inclusion of sound and video information, files grow rapidly in size, increasing their demands on the network. Even if most of the traffic follows the 80/20 rule, when the "20% application" is bandwidth-intensive, a routed internetwork can reach its limits. Emerging applications that can push the limits of routed internetworks include:

- Transfers of large graphics files, or combined voice/video/data information,
- Diskless workstations, which use the network as a virtual disk,
- X-terminals,
- Client-Server applications,
- Video servers, especially real-time video teleconferencing.

At some point, the solution to the problem isn't to simply add more routing capability or backbone bandwidth, the solution is to change the internetworking model.

(5) *Even traditional application operation can cause LAN congestion*. Even ignoring the newer, "exotic" applications, the natural growth of application use within an organization can cause workgroup LAN congestion. As more users are added to the workgroup, and existing users become more productive with existing applications, the demands placed on the network will increase over time. Files always grow in number and size, user populations increase, and more applications and application features are constantly being added to the mix. At some point, independent of the *inter*-network, the *intra*-network (the workgroup LAN) can become congested.

SYMPTOMS AND EFFECTS OF LAN CONGESTION

Until fairly recently, LAN congestion had never been a problem. LANs were historically designed around data rates in the millions of bits per second, far in excess of most computing devices' communications capabilities. However, advances in computing and communications controller technology have changed this scenario. Many devices today can (if given the opportunity) use the full channel capacity of a typical LAN. When many such devices share the channel, it is likely to cause congestion. Congestion is a statistical phenomenon; it occurs as a function of traffic patterns. LAN congestion manifests itself in a number of ways, as discussed below.[1]

Increased Network Delay

All LANs have a finite data carrying capacity. When presented with a short-term overload, the LAN distributes that load over time. Thus, when load is light, the average time from submission of a frame for transmission by the host until it is actually sent on the LAN will be short. When there is heavy instantaneous offered load, the average delay (known as *service time*) will increase. This has the effect of making the network appear "slower", since it takes longer to send the same amount of data under congestion conditions than it does when load is light.

It is difficult to directly measure service time (unless specially-instrumented driver software is configured for exactly this purpose). There are other, more observable metrics that can be used to infer congestion conditions.

Observable Parameters

Many parameters of LAN operation can be measured to assess network performance. Some parameters are automatically measured by standard controllers and host software; others typically require special network monitoring equipment, such as protocol analyzers or remote monitors (RMONs). Some of the important metrics include:

---

1 A more complete discussion of LAN congestion mechanisms can be found in:
Seifert, Rich, *The Effect of Ethernet Behavior on Networks using High-Performance Workstations and Servers*, Networks and Communications Technical Report, available by anonymous ftp:
ftp.netcom.com / pub / se / seifert / techrept13.ps

- *Channel Utilization*. Channel utilization refers to the percentage of time in which the channel is busy carrying data. It is directly related to the offered load.

  There are many variables to consider when trying to determine what constitutes an acceptable utilization, including: the number of stations on the LAN, application behavior, traffic patterns, frame length distribution, etc. Nonetheless, experience shows that for many common environments, including office automation LANs with tens of stations, the following utilization levels can be used as "rules of thumb" for determining when a LAN is approaching excessive load:

  - Utilization exceeds 10-20% averaged over an 8-hour work day, or
  - Utilization exceeds 20-30% averaged over the worst hour of the day, or
  - Utilization exceeds 50% averaged over the worst 15 minutes of the day

  Note that for very short-term periods (seconds, or even tens of seconds), network utilization may be nearly 100% without causing any problems. This might occur during a large file transfer between a pair of high-performance stations on an otherwise quiet network. Again, these are not hard-and-fast rules, and some application environments may operate well under heavier loads or fail at lighter levels.

- *Collision Counts.* An increase in the number of collisions on an Ethernet is also a (somewhat indirect) metric of offered load. High numbers of collisions are not necessarily indicative of a problem; Ethernet uses collision information to quickly redistribute the instantaneous offered load over the available time, maximizing channel utilization and application throughput.

  A major preoccupation with network administrators these days seems to be monitoring and worrying about the number of collisions seen on Ethernet networks. There is a great deal of folklore and voodoo concerning what is an "acceptable" collision rate, and when is the network "broken" or on the verge of collapse. In reality, collisions consume a very small percentage of available channel capacity, even under moderate-to-heavy offered load. As long as user performance and application throughput are acceptable, collision statistics can be generally ignored. Except in the most extreme of circumstances (all of which are observable through other, better metrics), the number of collisions seen on a network is an uninteresting and misleading statistic.

- *Application Performance Degradation*. LAN congestion will also manifest itself in lower application throughput. File transfers take longer and terminal sessions behave sluggishly when the LAN is congested. In some extreme circumstances, it is possible for an application to fail completely under heavy network load. Sessions may time out and disconnect, and applications or operating systems may actually crash, requiring a restart.

  It is important to realize that, while LAN congestion can cause application performance degradation, reduced application performance is not necessarily (or even usually) an indication of LAN congestion. Many factors contribute to application performance (e.g., CPU, memory and disk performance, number of users, etc.); the LAN is only one possible bottleneck.

- *User Dissatisfaction*. The ultimate manifestation of LAN congestion is user dissatisfaction. Regardless of all the collected statistics, if users are happy with the behavior of the system, then there really is no problem (at least, not now). If users are dissatisfied with the performance of the system, then all of the statistics indicating that the network is behaving correctly won't relieve their ire. User reaction is the most important metric of network performance.

  Again, user dissatisfaction with the performance of the network does not immediately indicate a congestion problem. The users' concept of "the network" includes the applications, servers, protocol stacks, internetworking devices, etc., and not just the underlying communications network.

TRADITIONAL SOLUTIONS TO LAN CONGESTION

Congestion means that there is more offered load than there is capacity of the LAN to carry it. All solutions to LAN congestion problems involve either increasing the LAN's capacity, or reducing the offered load.

Increasing LAN Capacity

If a 10 Mbps LAN is experiencing utilization of 50% averaged over a period of hours (typically an indication of congestion), the utilization would be on the order of 5% over the same period if the LAN capacity were increased to 100 Mbps. Thus, we can alleviate congestion by increasing the capacity of the LAN (as long as we don't simultaneously increase the offered load).

This "solves the problem", but has some undesirable side effects:

- *Increased cost*. Controllers, hubs, and internetworking devices designed for 100 Mbps operation (regardless of the particular technology) cost more than their 10 Mbps equivalents. While the relationship is not linear with data rate (i.e., it isn't 10 times the cost), it is still more expensive to use faster technologies.
- *Replacement of existing interfaces and equipment*. Not only is the higher speed equipment more expensive, there is usually an existing investment in the lower speed equipment already in place. So the higher cost is not really *instead of*, but *in addition to*, the lower speed equipment costs. Added to that is the expense of equipment replacement and installation, and disruption of network services during the changeover. Newly-installed networks (particularly high-speed LANs using less-mature products) often go through a period of unstable, problem-prone behavior, until the administrators become familiar with the equipment and its operation.
- *Replacement of media*. In many cases, a higher-speed LAN will require new cabling to be usable. Many existing cable plants were not designed to support high-speed LAN operation. Cables, connectors, patch panels and other wiring closet components may have to be replaced. For some technologies, copper wire may have to be replaced with fiber to be effective.

Segmenting the LAN to Reduce Channel Utilization

An alternative to increasing LAN capacity is to segment the LAN into multiple subnetworks, typically using routers. (See Figure 4.)

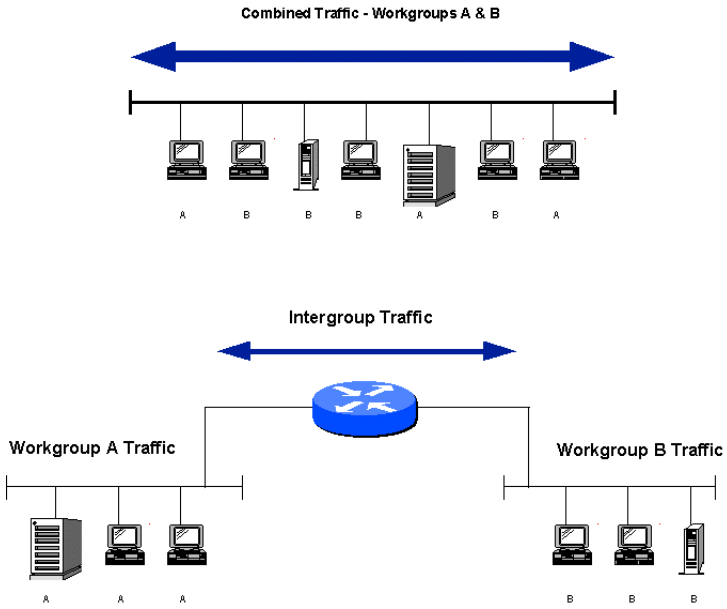

Figure 4: LAN Segmentation with Routers

Since the network now comprises multiple LANs for the same number of user devices (where there was only one LAN previously), we have effectively increased the available capacity of the aggregate system through segmentation. We have increased the available bandwidth on a per-user basis. On the other hand, we have also decreased the extent of the logical workgroup to the devices on each of the segmented LANs. The decrease in extent may be acceptable if, after segmentation the "80/20" rule still applies, i.e., traffic is localized within the new segments. Of course, the greater the granularity of the segmentation, the less likely will be the case that needed resources are local to the segment. In that case, segmentation will result in:

- *Increased internetwork traffic*. The router must handle all inter-segment traffic. As we segment more, we increase the amount of traffic that must pass through the router. Depending on the traffic patterns, we may have simply shifted the congestion problem from the LAN to the router, rather than actually solving the problem.
- *The need for additional servers or other resources*. In order to keep traffic local, it may be necessary to add servers to each local segment, in order to maintain the 80/20 distribution with the (now smaller) workgroup. This carries with it a cost penalty for the additional devices.

In many cases, none of the alternatives are attractive. They all carry some downside in exchange for alleviating the LAN congestion.

SWITCH CONCEPTS

A Local Area Network Switch provides an attractive alternative to the classical solutions for LAN congestion.

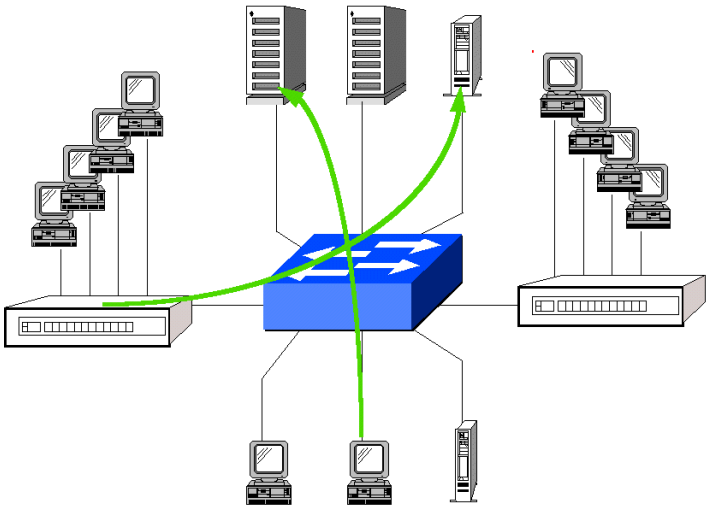

Figure 5: LAN Switch

A switch is a link-layer internetworking device that allows simultaneous frame exchange among large numbers of LANs and workstations. Conceptually similar to a massively-parallel LAN bridge, a switch allows information on any port to pass to any other port, simultaneously.[2] The switch inspects each incoming frame for the destination address of the target. It quickly determines the appropriate output port by consulting its internal address map. In the event that the output port is available, the switch *immediately* begins forwarding the frame to the destination, reducing the latency inherent in most bridge architectures that have to receive the entire frame before making a forwarding decision. This “cut-through” approach to switching greatly reduces the latency through the switch, and can improve application performance.

In the event that the target output port is not currently available, the switch buffers the incoming frame and forwards it when the output port becomes available. This gives the user all of the capability of both cut-through switching and store-and-forward bridging on a frame-by-frame basis, automatically.

Since the switch makes its forwarding decision based on link-layer addresses, it avoids the processing overhead inherent in routers, which must parse network layer header information, and make changes in the packet before forwarding it to the destination. It is the processing required to change packets before forwarding—typically adjusting a Time-to-Live field and recomputing header checksums—that either slows down a router or alternatively, demands more processing power (at higher cost) than a switch.

---

2 For a thorough treatment of the principles of operation of bridges, see:
Perlman, Radia, *Interconnections: Bridges and Routers*, ©1992 Addison-Wesley

Global Addressing

Switches offer a highly-efficient optimization for LAN internetworking. They take advantage of the fact that on a LAN, every device interface has a globally-unique 48 bit link address. By keeping track of the relative location of the devices in the interconnected LANs (i.e., which addresses are accessible through each port of the switch), the switch can quickly determine the appropriate port through which to forward traffic destined for that device. This can be done without the high overhead of using a network layer protocol, such as IP, for making this routing decision.

Unicast Operation and Learning

A switch can dynamically determine the mapping of link addresses to its ports. This is important, because:

- Devices may be moved from port to port,
- Interface hardware may change, changing the globally unique link address along with it,
- Topology changes in the internetwork may make devices *appear* to move relative to the switch ports.

Without some means for automatically determining the address-to-port mapping, switch administrators would be constantly fighting an uphill battle to keep the tables current and correct.

Fortunately, every frame sent by a device on a LAN carries the link address of the sender in the *Source Address* field of the frame. By listening to all frames, the switch can quickly determine which devices are connected to which ports. When devices move or topologies change, the switch can automatically update its internal tables through this *learning process*. Thus, there is never any issue or ambiguity regarding where a switch should forward unicast frames, since there is a 1:1:1 mapping among unicast addresses, device interfaces, and switch ports. (Unicast addresses are sometimes referred to as *physical addresses* or *individual addresses*. A unicast frame is a frame sent to a single destination device.)

Multicast Operation

LANs are unique in that they can offer the possibility of sending a single frame simultaneously to multiple receivers. *Multicast frames* are frames sent to a group of target destination devices. A multicast address refers to the group of devices that choose to receive the frame. Unlike the 1:1 mapping of unicast addresses, there is a 1:*n* mapping between multicast addresses and devices. In addition, a station never uses a multicast address as the source address in its transmitted frames. Thus, it is not possible to automatically learn the relative location of devices which are members of a particular multicast group.

Multicast groups are application specific; i.e., the set of applications running on a given device determines the set of multicast addresses that the device interface must recognize. A single device may listen to many multicast addresses, in addition to its unicast address. Multicast addresses are not “burned into” the device, as unicast addresses typically are.

The primary use of multicast today is to allow stations to learn about the services and resources available on the LAN. Servers can make their application services available by "advertising" their presence to a multicast address. Any station wishing to discover and use that application service needs simply to listen to the associated multicast. In this manner, a multicast group defines a "logical workgroup" for a particular application. Many such logical workgroups can exist on a single LAN, with each using a multicast (or more than one multicast) address to define the group. Figure 6 depicts how multicast addresses define logical groups on a single LAN.

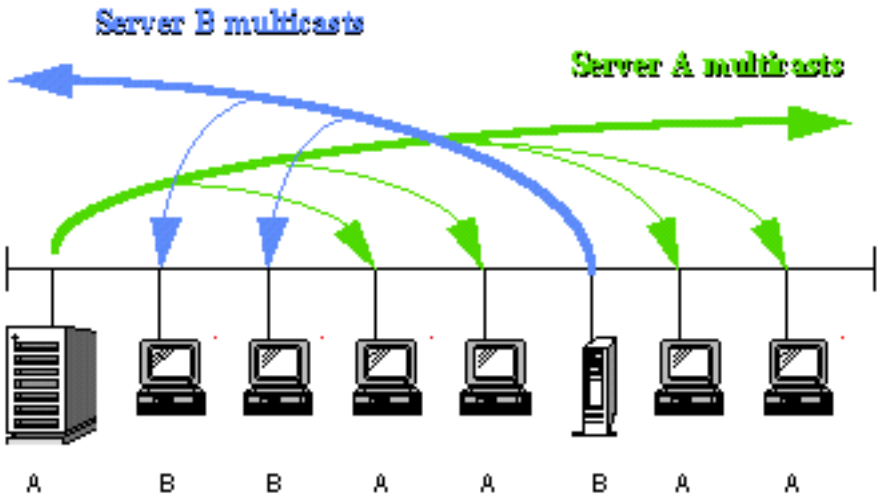

Figure 6: Multicast Address Mapping to Logical Workgroups

This concept is important for the creation of *virtual LANs*, discussed later.

Switches as an Alternative to Routers for LAN Segmentation

Remember that the primary reason for segmenting LANs was to reduce LAN congestion. Switches can offer a number of advantages over routers for this purpose:

Switches	Routers
Very low latency	More administrative control
Higher throughput (for given cost)	Better traffic isolation
Ease of administration and configuration (completely automatic)	Better fault isolation (firewall)

Advantages of Switches and Routers for LAN Segmentation

IMPLICATIONS OF SEGMENTATION USING SWITCHES

Microsegmentation

We have discussed the use of switches to segment LANs. Consider the scenario in Figure 7:

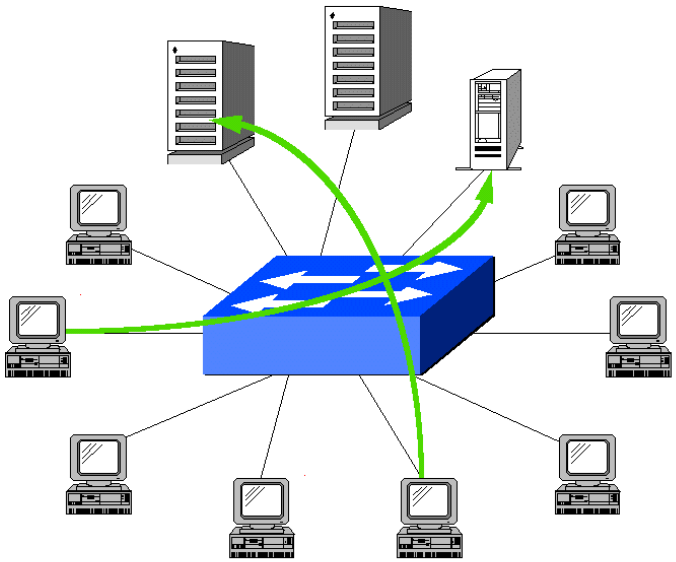

Figure 7: Microsegmentation

In this scenario, we have taken the concept of segmentation to the extreme. Each device has been isolated to its own LAN. Clearly no further segmentation is possible.

The result of this *microsegmentation* is that each device has been given the entire capacity of the LAN for its own use. In this manner, one of the primary causes of LAN congestion has been removed; there is never a problem with the aggregation of traffic from multiple stations causing a short-term heavy load condition, since there can be only one station offering load to the LAN at any given time.

As usual, there is no free lunch. In exchange for providing dedicated bandwidth to each station, we have shifted the congestion burden to the central switch. Note that microsegmentation is the ultimate violation of the 80/20 rule; 100% of LAN traffic must be internetworked when using this approach.

To the extent that the switch can be built to handle the total aggregate load of all of the attached devices, this "violation" of the 80/20 rule is not a problem. More important is the possibility of congestion due to traffic patterns causing load to converge on a given port. If many devices are all attempting to communicate with a single device (as show in Figure 8), then it is possible to have a congestion problem, even though every device has its own LAN.

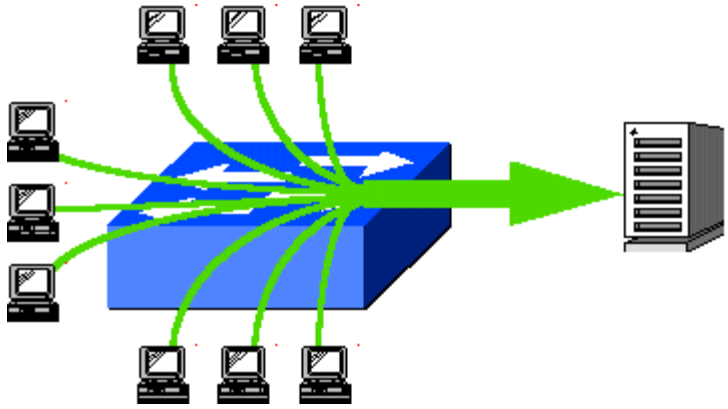

Figure 8: Traffic Congestion in a microsegmented LAN

There are two solutions to this problem:

- Provide a higher-bandwidth LAN to the devices where traffic converges (typically servers), or
- Provide a means for multiple ports to be connected to those same devices, effectively increasing the bandwidth without changing technologies.

Both have been implemented in practical switches.

Full Duplex Operation

Ethernet is normally a half-duplex communications system. While data can be transferred in any direction, at any given time a station is either transmitting or receiving, but not both. On the original physical media used with Ethernet (i.e., coaxial cable), this was the only communications possible, since the same wire was used for transmission and reception.

With current 10BASE-T (twisted pair Ethernet) technology, there are separate wire pairs for transmission and reception. However, since many devices are typically sharing the medium, we need a means to prevent multiple simultaneous transmissions. 10BASE-T uses the presence of received information during a transmission to indicate a *collision* to the transmitting stations, invoking the backoff and retransmission algorithm necessary for proper Ethernet operation.

In a microsegmented environment, we can be sure that there is only one device ever wishing to use a wire pair, since there is only one end station connected to each port of the switch. Only the attached station ever speaks to the switch (using the *Transmit* pair of the cable), and only the switch ever speaks to the attached station (using the *Receive* pair of the cable). There is never any contention for the use of the medium, as in a standard Ethernet environment.

With the possibility of contention removed, we no longer have any real need for the collision detection and backoff function. They can be eliminated, and both the station and the switch allowed to transmit at will, in both directions simultaneously. This is depicted in Figure 9.

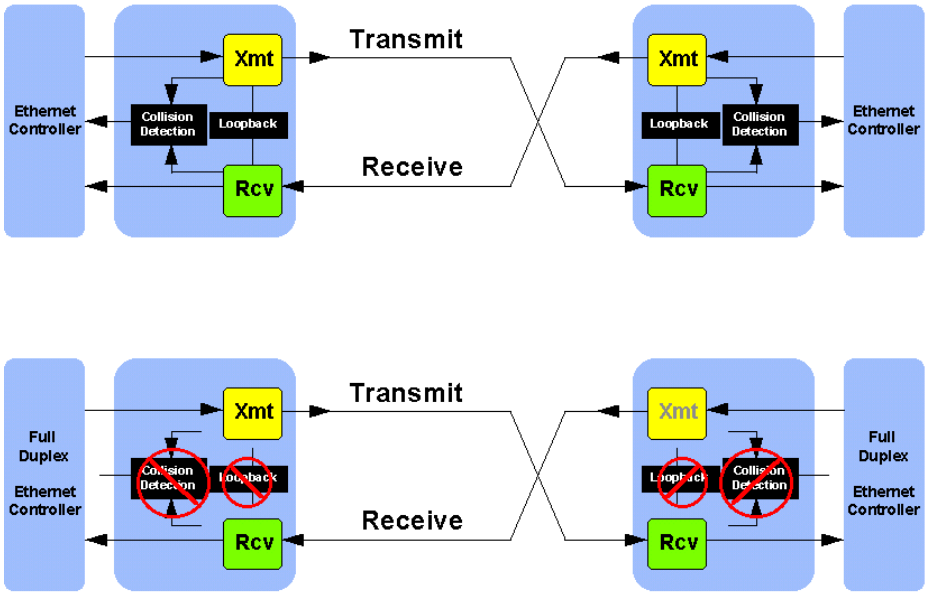

Figure 9: Half-duplex and Full-duplex Ethernet

The effective data transfer rate has been doubled, since we can now transmit the full channel data rate in both directions. This is true regardless of the half-duplex data rate (i.e., full-duplex signaling can be used on both 10BASE-T and 100BASE-T). The only requirement is that:

- *Separate wire pairs are used for transmit and receive; no pairs serve dual functions.* 10BASE-T, 100BASE-TX, 100BASE-FX, Token Ring and CDDI meet this criterion. 100BASE-T4 and 100VG-AnyLAN using Quartet Signaling™ do not.
- *Only one device is connected to each port of the switch.* This eliminates any contention for the use of the channel in either direction.
- *The central hub is a switch.* Full-duplex operation is not possible with traditional repeated hubs.

It is important to realize that application behavior will determine whether a station can really take advantage of this increased capacity. Typical applications running on LAN-attached stations today do not use bandwidth symmetrically. For example, file transfers (which typically can benefit from increased bandwidth) are asymmetrical in nature; bulk data is transferred in one direction, with short acknowledgments being returned in the reverse direction. A station which is only doing file transfers will not benefit from the doubling of bandwidth offered by full-duplex switching.

Where the benefits of full-duplex make sense is when there are multiple applications simultaneously using the network on a single device. This can be a multitasking workstation (e.g., UNIX or Macintosh, but not DOS), or a server. Typically servers are handling network traffic to and from multiple workstations simultaneously. While any given workstation is using the network asymmetrically, the server can take advantage of full-duplex operation to simultaneously handle transfers in from one station, and out to another.

In the future, applications that use high-capacity LANs symmetrically (e.g., video teleconferencing) may become popular. Such applications can greatly benefit from full-duplex operation, even in a single-tasking workstation.

Virtual LANs

Earlier we discussed the use of multicast addresses to advertise services and resources to stations on the network. In this manner, a multicast group defines a logical grouping of devices implementing a particular application. By default, a switch will forward all multicast traffic onto all ports (except the port on which the frame arrived). This provides full multicast connectivity among all stations on the switched LAN.

However, it is possible to have the switch *isolate* multicast propagation to specific ports, under administrative control. As shown in Figure 10, this can create logically separated workgroups, or *virtual LANs,* within the switched infrastructure.

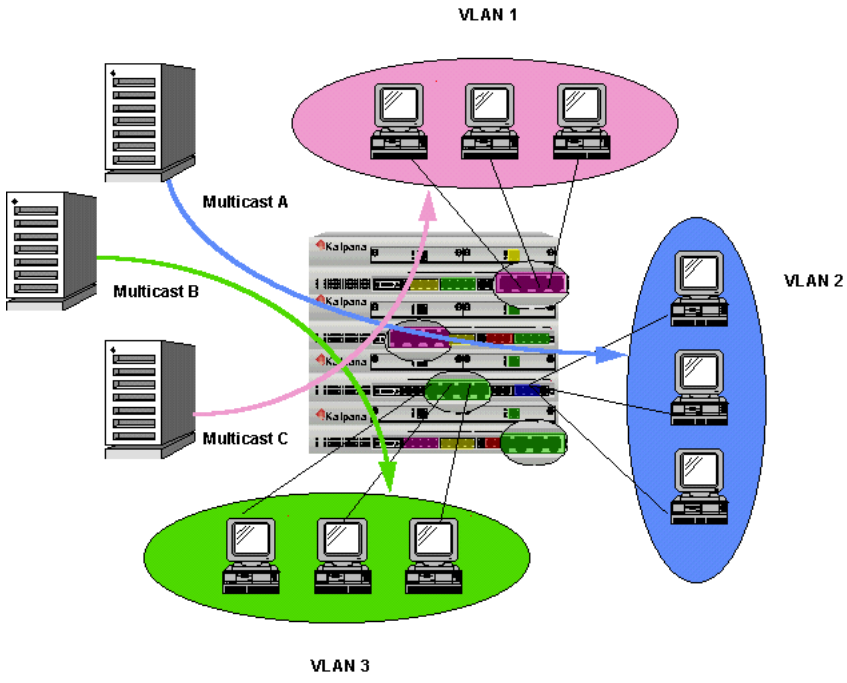

Figure 10: Creating Virtual LANs through Multicast Filtering

The switch will automatically isolate unicast traffic, through the learning algorithm and normal switch operation. By setting up specific multicast filter tables in the switch, we can create true virtual LANs, where there is full connectivity, yet a logical separation of devices. By associating the multicast filters with the device, we can allow devices to move arbitrarily within the switched internetwork, yet still appear to be a member of the same logical workgroups. Virtual LANs thus allow us to separate the concepts of physical and logical connectivity.

The advantages of this approach are enormous. Currently, a workgroup is defined by those devices plugged into a given hub. Using virtual LAN technology, we define a workgroup not by where the device is plugged in, but by the multicast filters in the switch, associated with the applications running on that device. If a device moves within the switched internetwork, we can move the multicast filters automatically, maintaining logical connectivity without having to physically intervene.

Currently, multicast filters must be defined by a network administrator, under manual control. Work is ongoing in the industry to develop standards for automatic registration (and authentication) of multicast filters in a switch by the attached stations. This will further enhance the power of virtual LAN technology.

SWITCHING ISSUES

Migration to higher speed station connections

Up to now, we have generally assumed that all of the switch's connections used the same technology (e.g., 10BASE-T Ethernet). This is not strictly necessary. Since a switch only depends on the use of globally-unique link addresses for its operation, any LAN that uses this address structure is a candidate for a switch port. There is no requirement that all ports on a given switch use the same technology. This is depicted in Figure 11.

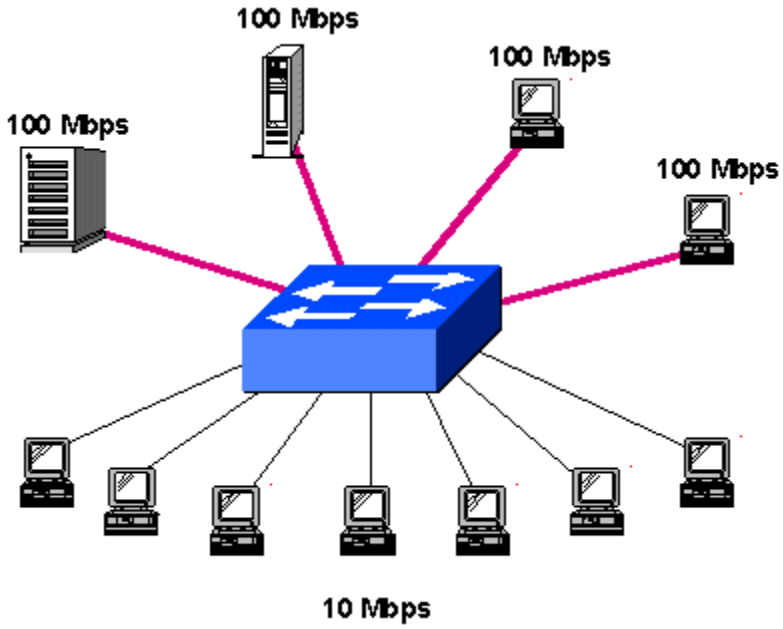

Figure 11: Mixing Data Rates on a Single Switch

In theory, then, one could mix Ethernet, Token Ring and FDDI ports on the same switch. In practice, this would eliminate some of the advantages of switching. Since each of these technologies uses a different frame format, frames would have to be "re-built" before forwarding between ports using dissimilar technologies. Link layer checksums would have to be recomputed as well. This added processing burden would reduce the performance (or increase the cost) of a switch.

However, it is perfectly reasonable to use the same technology *at different data rates* on different ports. Thus, a switch could easily move frames between Ethernet ports running at 10 Mbps and 100 Mbps (100BASE-T, Fast Ethernet). Since there is no change in the link layer frame format, there is no need to recompute checksums or otherwise rebuild the frames before forwarding.

In a microsegmented environment we can gain even more benefits from higher speed operation. The change from 10 Mbps to 100 Mbps operation can be done on a *station-by-station basis*, i.e., we can choose to migrate individual stations to high-speed operation, independent of all other stations. We may choose to upgrade the servers first, then perhaps some high-performance workstations, yet leave the bulk of the "casual" users at 10 Mbps. Everyone will benefit from this migration, yet costs can be contained. The use of switching technology allows the network to move to high-speed operation without a major overhaul, conserving the investment in existing equipment.

Because of the differences in speed, it is not possible to operate in a cut-through manner when switching between dissimilar speed ports, however all of the other benefits of switching still accrue, and cut-through operation is maintained among ports operating at the same speed (either 10 Mbps or 100 Mbps).

It is also possible to achieve higher-speed operation without using high-speed LAN equipment, i.e., without changing any controllers or switch ports. Multiple 10 Mbps ports can be aggregated using multiple, (low cost) standard interface controllers at the station, as shown in Figure 12.

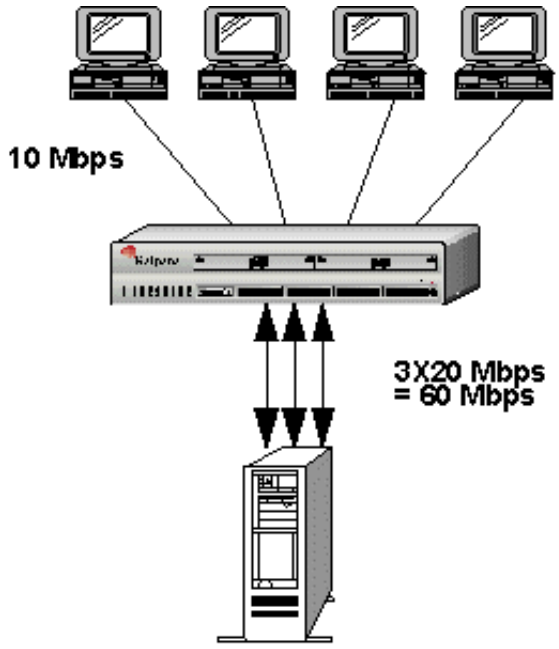

Figure 12: Aggregating Port Capacity

Higher aggregate data rates can be achieved by combining multiple ports. While not as effective as using a 100 Mbps solution, up to 30 Mbps capacity (60 Mbps in full-duplex mode) can be achieved with minimal investment in equipment.

Whether you use 100 Mbps connections or multiple 10 Mbps pipes, by providing higher capacity to those devices where network traffic converges (typically servers), the LAN congestion problem can be greatly alleviated. Remember that in a switched environment, unless traffic patterns are uniform across ports, there can be significant congestion within the switch. By providing a "thicker pipe" to those devices that source or sink large amounts of traffic, we can reduce or eliminate this congestion point.

High Speed Switch Interconnections

High speed links not only make sense for high-performance network stations, they are the obvious choice for interconnection of switches in a switch hierarchy. An ideal approach to modular switch design would have some number of 10 Mbps ports (for typical user workstation connections), and a smaller number of high-speed ports. The high-speed ports could be used for server connections (as discussed above), or for inter-switch connections in a *stackable* switching hub, as shown in Figure 13.

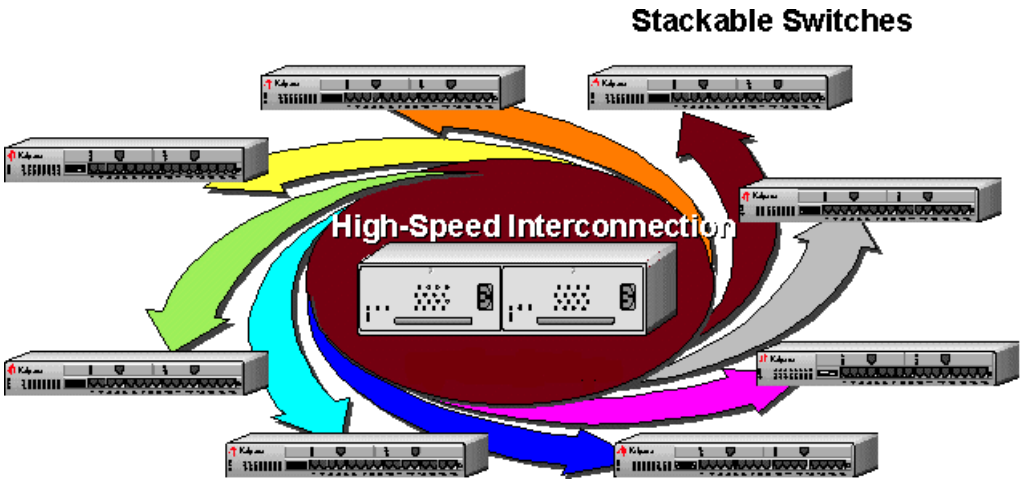

Figure 13: Stackable Switching Hub Configuration

The technology used for the inter-switch connection could take many forms:

- *100BASE-T.* This is an obvious choice, since it is appropriate both for high-speed server connections and within-a-rack stackable hub connections. 100BASE-T has a distance limitation of 100 meters for any link, so it is not appropriate for campus or WAN switch connections.
- *Proprietary connections.* For a switch-to-switch connection, a proprietary solution may be appropriate. Higher capacities than 100 Mbps may be achieved at reasonable cost by leveraging off the implementation-specifics of a particular switch design. While this approach does not allow interconnection of switches from different vendors, this is rarely a problem; in practice all of the switches will likely be from the same supplier (at least within a given stackable hub).
- *FDDI.* FDDI provides 100 Mbps data rate, but can cover distances up to 100 km (ring circumference). This is ideal for campus or metropolitan area switch interconnection, as shown in Figure 14:

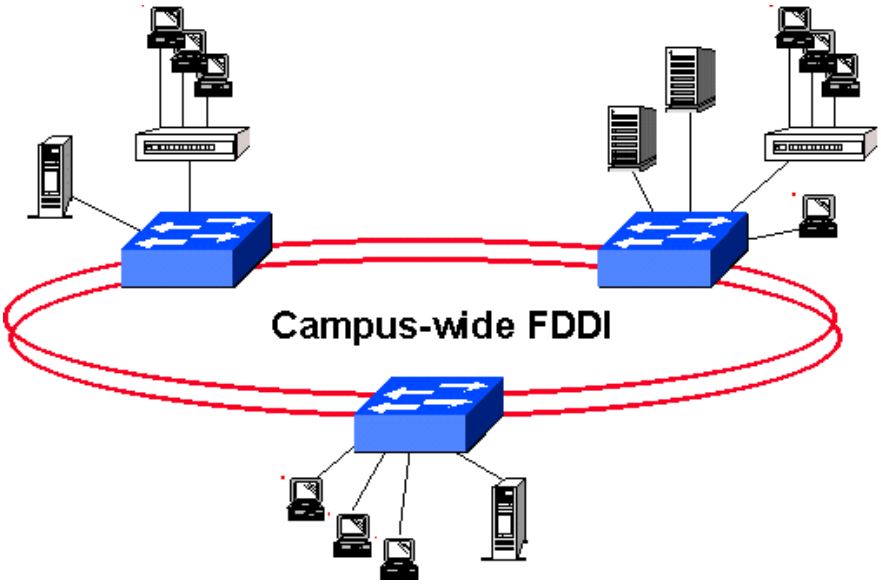

Figure 14: FDDI-connected Campus Switch

- *ATM*. ATM offers the possibility of high-speed switch connection across even larger distances than FDDI. As OC-3 (155 Mbps) ATM WANs become available, this technology will allow worldwide interconnection of switches in an enterprise network.

Integrating the Switched Workgroup Cluster to the Routed Enterprise Network

Switches are an effective means of interconnecting LANs. While they offer higher performance at a given cost than routers, this does not imply that routers are never appropriate for internetworking. In exchange for their higher cost, routers offer a higher level of service. As discussed earlier, routers can provide security, network management, firewall, and route optimization features not available in switches. Even though the capabilities of switches and routers differ, there is no need for a user to make an absolute choice between switching and routing. It is possible and desirable to take advantage of the benefits of both technologies, *at the appropriate level* in the enterprise network hierarchy.

The benefits of switching accrue primarily to workgroups. Switches offer high-speed, low latency connections among ports. Microsegmentation (dedicated bandwidth to both workstations and servers), full-duplex operation, Virtual LAN creation and above all, ease of configuration and administration are all desirable characteristics in workgroup LANs.

The benefits of routers are most applicable in the connections *between* workgroups or clusters of workgroups. Security, route control, management and firewalling are important across organizational boundaries, or where expensive WAN links must be carefully managed. Figure 15 shows how switches and routers can be used together to get the best advantages of both technologies in an enterprise network.

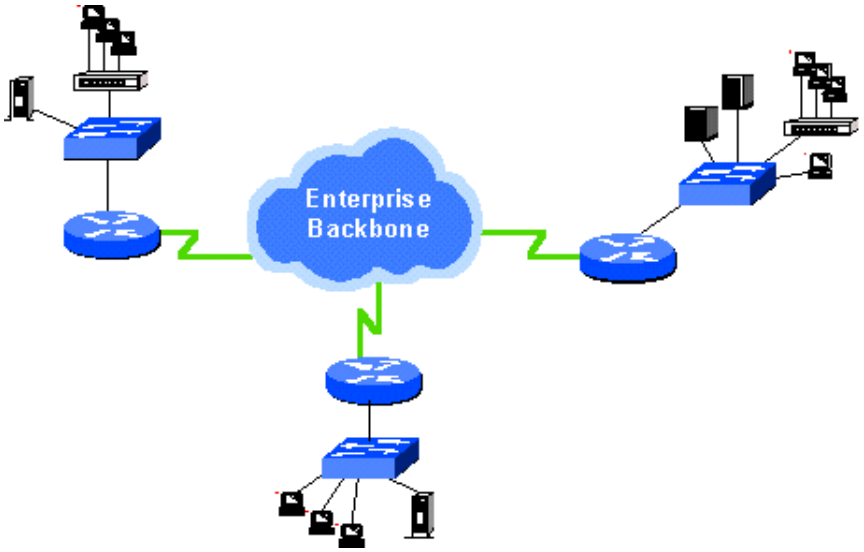

Figure 15: Switching and Routing in an Enterprise Network

MODELS OF SWITCH USAGE AND EVOLUTION FROM A SHARED LAN ENVIRONMENT

Most LAN environments today use shared LAN hub technology, with workgroups interconnected by routing to a campus backbone (either distributed or switched), as discussed earlier. As applications evolve and the user demands increase, internetwork traffic will increase. Congestion in the internetwork can become a problem, which can be addressed by switching technology.

The Switch of Hubs

Figure 16 depicts the first stage in this evolution, called a "Switch of Hubs". In this model, a switch is used as a collapsed backbone to interconnect workgroup LANs.

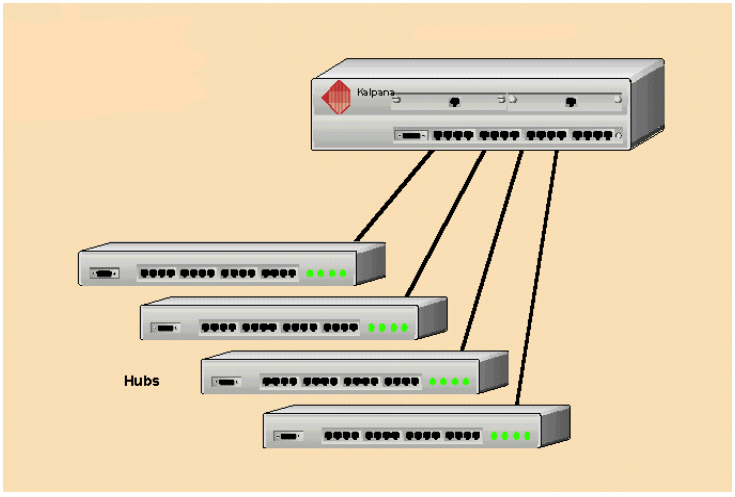

Figure 16: Switch of Hubs

This effectively eliminates congestion concerns with inter-group traffic, especially relative to using a shared hub as the backbone. Optimum advantage is obtained when the traffic patterns are relatively uniform among workgroups, i.e., there is no single workgroup that is the source or sink of a vast majority of traffic. Uniformity of internetwork traffic is the most common case, as workgroups tend to be autonomous, and there is rarely a strict hierarchy that reflects itself in the internetwork traffic patterns.

The Switch of Hubs is a good solution to the workgroup interconnection problem, and offers the following additional characteristics:

- No changes are required to the network interface hardware (NICs) in any of the user or server stations,
- A single switch can often support all of the workgroups within an organization,
- There is minimal disruption to the network; only the second-tier hub needs to be changed to a switch to get the performance advantage.

It is important that the switch used for a Switch of Hubs be capable of supporting large numbers of end stations per switch port. There may be tens or hundreds of stations connected to the shared hub attached to that port; there may even be a hierarchy (multi-level) of shared hubs, invisible to the switch.

The Switch of Hubs is a common first step in LAN evolution from a flat, shared environment to a hierarchical switched infrastructure.

The Switch of Servers

Most LAN traffic is directed either to or from a network server. In many cases, the limiting factor on application performance may be the capacity of the communications channel to the server, rather than the processing performance of the server. This is especially true when very high-performance servers ("superservers") are deployed, or when specific applications make large bandwidth demands on the LAN (e.g., image or bulk data transfer). In such cases, overall performance can be improved by using a "Switch of Servers", as depicted in Figure 17.

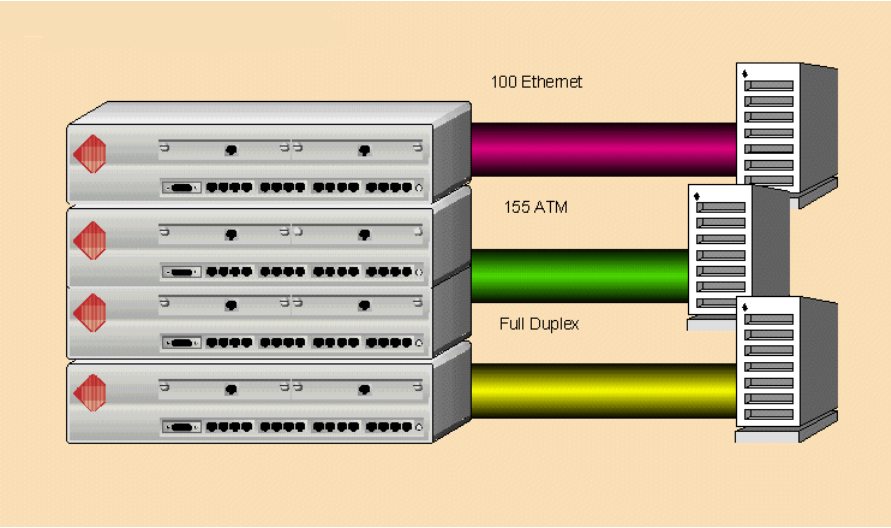

Figure 17: Switch of Servers

Here, a dedicated connection is provided for the servers to alleviate network congestion on this path. In the simplest case, the connection may be a standard 10 Mbps Ethernet; the advantage is gained by not having to share the channel with other stations and servers in the network. If greater throughput is required, the same approach can be used with:

- Full-duplex 10 Mbps Ethernet, offering up to 20 Mbps total capacity,
- Fast Ethernet, with up to 100 Mbps (half-duplex) or 200 Mbps (full-duplex) capacity ,
- ATM, with up to 310 Mbps capacity (i.e., 155 Mbps full-duplex).

An intermediate approach is also possible, using multiple 10 Mbps Ethernet ports to behave as if they were a single, higher-speed port from the server's perspective. This is depicted in Figure 12. This allows additional channel capacity to be added, without having to implement any of the newer, less mature technologies such as ATM. It also allows the user to take advantage of low-cost 10 Mbps NICs that may already be available on-site. Depending on the Network Operating System (NOS) in use, (e.g., NetWare), a special device driver (e.g., implemented as a NetWare Loadable Module, or NLM) may be required to use this option.

In addition to server connections, many application environments can benefit from providing increased channel capacity from the workgroup cluster to the enterprise backbone. High-performance routers may be treated as "internetwork servers" in this regard, and likewise provided with dedicated, switched connections, as depicted in Figure 17.

In order to deploy a Switch of Servers, it will generally be necessary to replace the NIC within the servers (and / or high-performance routers) from conventional 10 Mbps Ethernet to either Full-duplex Ethernet, Fast Ethernet, or ATM. However, only these connections need to be upgraded; no changes are required to the end stations, which constitute the vast majority of devices in the network. For a relatively small investment in server and switch hardware, a big improvement can be achieved in user throughput in a client-server environment.

The Switch of Desktops

As application and user demands grow, ultimately the capacity of the LAN becomes a bottleneck not only for servers and clusters of users, but for individual user devices as well. Network-intensive applications drive the need for more and more devices to have LAN bandwidth dedicated to an individual device. The natural extension of this philosophy leads to microsegmentation, as discussed earlier. In this scenario, the model is of a "Switch of Desktops", as depicted in Figure 18.

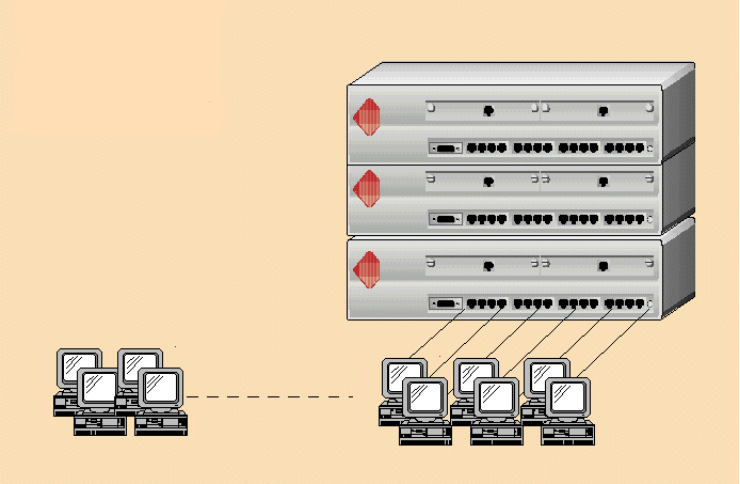

Figure 18: Switch of Desktops

This model represents a significant change in the paradigm of network interconnection relative to today's shared LANs. Every device has a LAN "pipe" dedicated to itself; the bandwidth of the pipe can be tailored to the needs of the station applications. If the end station is using standard 10 Mbps Ethernet, there is no additional investment required for NICs in the stations; only a shared hub needs to be replaced with a switching hub to reap the benefits. Depending on user requirements, stations can migrate to full-duplex, or 100 Mbps connections *one at a time*, without requiring wholesale changes to the network (and incurring large costs for overhaul).

In a Switch of Desktops, the switch now becomes the focal point of the LAN, and a number of critical concerns are raised:

- Are there a sufficient number of device ports available?
- Can the number of switch ports be expanded to meet increasing numbers of users?
- Can device ports be configured for 10 Mbps, 100 Mbps or higher data rates?
- Does the switch have adequate aggregate capacity to handle the combined load of all of the end stations?
- Can the switch be managed in a simple, consistent manner, regardless of station or configurations?

Migration is almost trivial from a shared LAN to a Switch of Desktops, especially when the majority of the devices are using 10 Mbps Ethernet. Devices can be moved, one at a time, from ports on a shared hub (current configuration) to ports on the switch simply by changing a patch panel connection in the wiring closet. In minutes, and with no change to the NIC or the software in the end station, the user can get the benefits of switching and dedicated bandwidth. During the migration, stations using shared and switched connections intercommunicate transparently, with no interoperability concerns. Low-performance or lightly used stations may be left on shared hub connections indefinitely, getting maximum benefit from the prior investment in shared LAN hub technology.

For stations in the switched infrastructure, the power of Virtual Networking becomes possible. Logical workgroups can be dynamically created and modified, regardless of the physical connections; a virtual workgroup can comprise members whose devices are connected to physically separate switches.

The Switch of Wiring Closets

As a last step in migration from a shared LAN to a switching architecture, we can use a switch as a collapsed backbone, interconnecting other switches being used for workgroup concentration. This is referred to as a "Switch of Wiring Closets", and is depicted in Figure 19. This is really a compound model, combining a Switch of Hubs (the backbone), where each hub is now a Switch of Desktops, rather than a shared LAN.

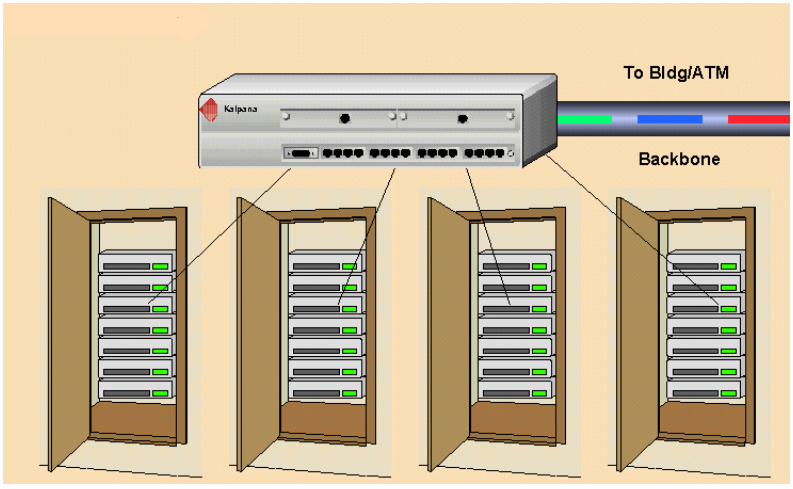

Figure 19: Switch of Wiring Closets

Connections from the wiring closet switches to the collapsed backbone can use Full-duplex Ethernet in all cases, including 100 Mbps Fast Ethernet for connections to truly bandwidth-intensive workgroups. The key considerations in the choice of a switch for this application include:

- Ability to grow from a small to a moderate number of ports (where each port connects to a wiring closet, as opposed to an individual device),
- High internal capacity, to eliminate backbone bottlenecks,
- High-speed ports for connection to high-performance routers on the enterprise backbone,

SUMMARY

In this paper, we have seen how switching technology provides a flexible, scalable solution to the problems inherent in current routed workgroup interconnection. In addition, switching opens the door to more powerful and flexible internetworking architectures, including microsegmentation, full-duplex operation, and virtual LANs. We have also analyzed four common models of use for switching hubs, and shown evolution paths from a shared to a switched LAN environment.

The author wishes to thank cisco Systems, Inc. for their support in the development of this report, and for the use of some of the figures it contains.